Listening. Learning. Leading.®

Research Report

ETS RR–14-18

A Generalization of Pythagoras's Theorem and Application to Explanations of Variance Contributions in Linear Models

James E. Carlson

December 2014

ETS Research Report Series

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Report series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Report series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

ETS Research Report Series ISSN 2330-8516

RESEARCH REPORT

A Generalization of Pythagoras's Theorem and Application to Explanations of Variance Contributions in Linear Models

James E. Carlson

Educational Testing Service, Princeton, NJ

Many aspects of the geometry of linear statistical models and least squares estimation are well known. Discussions of the geometry may be found in many sources. Some aspects of the geometry relating to the partitioning of variation that can be explained using a little-known theorem of Pappus and have not been discussed previously are the topic of this report. I discuss, using the theorem, how geometric explanation helps us understand issues relating to contributions of independent variables to explanation of variance in a dependent variable. A particular concern that the theorem helps explain involves nonorthogonal linear models including correlated regressors and analysis of variance.

Keywords Linear models; variance partitioning; vector geometry

doi:10.1002/ets2.12018

In many instances in statistical analyses as well as in interpretation of certain concepts in measurement, researchers include computation and discussion of the variance contribution of predictors in a regression analysis, or contributions of subscores to a composite score. In the author's view, there is a great deal of misunderstanding and misinformation in the literature about how we can legitimately break down explained variance into components in regression analysis (see, e.g., Carlson, 1968; Chase, 1960; Creager, 1969, 1971; Creager & Boruch, 1969; Creager & Valentine, 1962; Guilford, 1965; Hazewinkel, 1963; Mood, 1971; Newton & Spurrell, 1967; Pugh, 1968; Wisler, 1968) and about how different analysis of variance procedures are testing different hypothesis in nonorthogonal designs (see, e.g., Bancroft, 1968; Carlson & Timm, 1974; Searle, 1971; Speed, 1969; Timm & Carlson, 1975; Yates, 1934). In this work, based on Carlson and Carlson (2008a), I present geometric explanations related to this issue and show the problems that can occur. In a companion work (Carlson, in press), based on Carlson and Carlson (2008b), I discussed the issues associated with contributions of subscores to overall test scores.

I start with a review of the geometric background required to follow later discussion of the issues. Some of the basic algebra and geometry is covered in the Appendix. Readers having familiarity with this background can easily skip this review and the Appendix.

Geometric Background

As shown by several writers (Draper & Smith, 1966; Wickens, 1995; Wonnacott & Wonnacott, 1973), two different geometries may be used to show relationships between variables. The most commonly used geometry shows the variables as orthogonal (at right angles) axes and values of individuals on the variables as points in the Euclidean space defined by the axes. This is the *geometry in the variable space*. In this geometry, it is common to discuss the line of best fit by the ordinary least squares (OLS) criterion, the slope of that line, and the clustering of points about it in relation to the correlation and regression between the two variables. Variance is a measure of the spread of points parallel to the axis representing each variable. The alternative geometry treats each person as an axis (with orthogonal axes) and represents the variables as points in the Euclidean space. It is referred to as the *geometry in the person space*. The orthogonality of the axes is clearly an accurate indication that in linear models the different individuals are assumed independent of one another; orthogonality is equivalent to zero correlation.[1] In the geometry in the variable space, on the other hand, the axes are shown as orthogonal when in fact the variables are usually correlated — somewhat of a misrepresentation although the correlation is clearly indicated by the pattern of points.

Corresponding author: J. Carlson, E-mail: jcarlson@ets.org

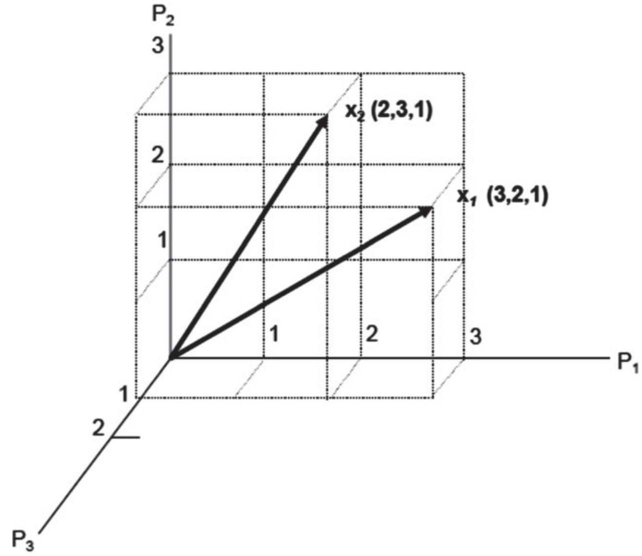

Figure 1 Geometry of two variables shown as vectors in a three-dimensional person space.

In the person-space geometry, the variables are often displayed graphically as vectors (directed line segments) drawn from the origin to the points, as shown in Figure 1 and in more detail in the Appendix. I use the convention of representing two variables as X_1 and X_2, and their vectors of N sample values as $\mathbf{X}_1$ and $\mathbf{X}_2$. In Figure 1, I illustrate with three persons (displayed as axes P_1, P_2, and P_3) with Person 1 having scores of 3 and 2 on variables X_1 and X_2, respectively; Person 2 having scores of 2 and 3; and Person 3 having scores of 1 and 1. With a sample of N persons in a study, this space is N-dimensional; but with only two independent variables, as shown in Figure 1, we only need a two-dimensional subspace of that N-dimensional space to represent the data. This fact, illustrated below, greatly simplifies the geometric explanation.

In the Appendix, I summarize relevant material covered in detail by the aforementioned authors who show that, in this geometry, the length of each vector is proportional to that variable's variability (sum of squares or standard deviation, depending on the metric) and the sum of two variables is a vector in the same plane as those variables; the latter can be found by appending one component vector onto the end of the other.

Partitioning Variation

Partitioning of variance in linear models can be explained geometrically using Pythagoras's theorem in the case of orthogonal (uncorrelated) independent variables (Draper & Smith, 1966; Kendall & Stuart, 1967; Wickens, 1995; Wonnacott & Wonnacott, 1973). The orthogonal case includes factorial analysis of variance with equal subclass numbers and uncorrelated predictors in regression analysis (which rarely occurs in practice). All of the derivation of analytical methods used with orthogonal linear models, including univariate and multivariate analysis of variance and OLS regression analysis, except the distributional theory of test statistics and interval estimation, can be accomplished using nothing more complex than Pythagoras's theorem and can be demonstrated using Euclidean geometry. In a later section, I demonstrate the use of Pythagoras's theorem and a theorem of Pappus, generalizing Pythagoras's theorem to non-right triangles, to clarify the meanings of several definitions of contributions to variance explained in linear models.

To discuss the main topic of this work, a linear regression model with one dependent variate, Y, regressed onto two regressors (independent variables or predictors), X_1 and X_2, will be used. In this discussion, I use a deviation score metric in which values on all three variables are expressed as deviations from their means. The issues involved readily generalize to the case of more than two regressors. The related geometry that will be used is presented in the Appendix. Using the deviation score metric, Figure A3 in the Appendix shows the regression for N sample observations as the perpendicular projection of the vector, $\mathbf{y}$, of values on Y, onto the plane spanned by vectors, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, of values on X_1 and X_2, respectively. Note that $\mathbf{y}$ lies outside the plane unless it is linearly dependent on $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ (in which case it would be perfectly predictable, without error). It is well known that the proportion of variation (i.e., sum of squares) in the dependent variate that is

accounted for by or predictable from the regressors in the sample data is the coefficient of determination, R^2. I begin by describing two different methods purported to partition variance explained in regression models into parts attributable to the different regressors.

The Chase-Engelhart-Guilford (CEG) Method

In the Chase-Engelhart-Guilford (CEG) method (Chase, 1960; Engelhart, 1936; Guilford, 1965), Equation A6 for the R^2 is used in attempts to partition this proportion of explained variance into portions uniquely associated with each regressor and jointly with pairs of regressors. The methodology uses a standard score metric (mean zero, variance one), which is simply a transformation of the deviation scores by dividing values on each variable by their standard deviations. In this metric, I use z_y to represent the vector of standardized values on the dependent variate and z_j to represent those on regressor j.

The statistics defining the CEG unique contributions of the p regressors to the prediction, using the standardized metric, are the p terms of the first expression on the right of Equation A6,

$$V_j = \left(b_j^*\right)^2 \quad \left(j = 1, 2, \ldots, p\right). \tag{1}$$

The b_j^* are the regression coefficients in the standardized metric. Similarly the joint contributions in this method are defined from the final term on the right of A6 as:

$$J_{jj\prime} = 2b_j^* b_{j\prime}^* r_{jj\prime}, \quad \left(j, j\prime = 1, 2, \ldots, p\,; j \neq j\prime\right), \tag{2}$$

where $r_{jj\prime}$ is the correlation between $\mathbf{x}_j$ and $\mathbf{x}_{j\prime}$. Guilford (1965, pp. 399–400) referred to the V_j as direct contributions of the X_j. Statistics defining the total contributions (Chase, 1960) in the CEG method are taken from Equation A7 as:

$$T_j = b_j^* r_{yj} \quad \left(j = 1, 2, \ldots, p\right). \tag{3}$$

Bock (1975) also discussed the terms in Equation 3, stating that "Only when the X variables are uncorrelated in the sample . . . are these terms nonnegative and do they represent proportions of predictable variation" (p. 380). He stated "regrettably, there are many erroneous interpretations of R^2 in the literature. Perhaps the worst is the identification of the jth term [of Equation 3] as the proportion of variance attributable to the jth predictor." Guilford (p. 400) stated that, to be interpreted as variance contributions, the terms in Equation 3 "must all be positive" (p. 400), but Bock is more correct in including zero (although a zero term would indicate zero contribution, so it is a moot point). Carlson (1968) showed that an alternative definition of the variance contribution of predictor j, proposed by Richardson (1941), is equivalent to Chase's T_j so it will not be discussed further here. Guilford referred to the T_j as *direct plus indirect* contributions. He also referred to the difference,

$$T_j - V_j = b_j^* r_{yj} - \left(b_j^*\right)^2, \tag{4}$$

as the *indirect* contributions. Note that, using Equations 4 and A6, in the two predictor case,

$$\left[b_j^* r_{yj} - \left(b_j^*\right)^2\right] + \left[b_{j\prime}^* r_{yj\prime} - \left(b_{j\prime}^*\right)^2\right] = 2b_j^* b_{j\prime}^* r_{jj\prime}.$$

So Guilford's indirect contributions of predictors X_j and $X_{j'}$ sum to the Chase-Engelhart joint contribution, $J_{jj'}$ of these two regressors. Similarly the sum of all p of Guilford's indirect contribution terms in Equation 4 equals the sum of all $p(p$-1) Chase-Engelhart joint contributions. Guilford does not include joint contributions in his discussion. Bock (1975, pp. 380–381) cited Wright's method of path coefficients as "a correct, but less straightforward, interpretation" and, in discussing this method, described the terms of Equations 1 and 2 as direct and indirect contributions, respectively. Hence, we see that different authors use different terms in discussing the same statistics as variance contributions.

The Creager-Valentine (CV) Method

The Creager-Valentine (CV) method (Creager, 1969, 1971; Creager & Valentine, 1962; Hazewinkel, 1963; Mood, 1971; Newton & Spurrell, 1967; Wisler, 1968) is another method for partitioning the variation. The method derives from fitting

several models containing different combinations of regressors; some of these authors refer to the method as *commonality analysis*. Differences between the R^2 values from the different models are used to define the unique and joint contributions of regressors. Note that because R is a correlation coefficient, it is the same in any of the metrics I am using. Hence, the discussion in this section requires no specification of the metric.

The CV procedures are mathematically identical to the method of fitting constants commonly used for many years in the analysis of variance for nonorthogonal designs (Bancroft, 1968; Searle, 1971; Yates, 1934) and the method of part correlations (Creager & Boruch, 1969; Pugh, 1968). They can also be shown to be equivalent to regressing Y onto orthogonal component variables derived by different Cholesky factorizations of the predictor correlation matrix. Using these procedures with two regressors, the unique contribution statistics, here denoted by U_1 and U_2 for X_1 and X_2, respectively, are:

$$U_1 = R^2_{y.12} - r^2_{y2}, \text{ and}$$
$$U_2 = R^2_{y.12} - r^2_{y1}, \tag{5}$$

where $R^2_{y.12}$ is the proportion of variance accounted for by the two-regressor model. Because r^2_{y2} is the proportion of variance accounted for by regressing Y onto X_2 in a one-predictor regression, U_1 in Equation 5 is the increase in the proportion accounted for when X_1 is added to form the two-regressor model. A similar interpretation can be made for U_2 as the increase when X_2 is added to a one-predictor regression involving only X_1. In the CV procedure the joint contribution of X_1 and X_2, here denoted as K_{12}, is attributed to the remainder,

$$K_{12} = R^2_{y.12} - U_1 - U_2. \tag{6}$$

All of the previous equations may be generalized to the case of more than two predictors, in which case there are definitions of unique and joint contributions of each predictor and each pair of predictors, respectively. The generalizations of Equation 5 are:

$$U_j = R^2_{y.1-p} - R^2_{y.1-p,\tilde{j}},$$

where $R^2_{y.1-p}$ is the proportion of variance accounted for by the p variable regression and $R^2_{y.1-p,\tilde{j}}$ the proportion accounted for by the $p-1$ variable regression with variable X_j deleted from the model.

Carlson (1968) showed that the relationship between the CEG V_j and the CV U_j is:

$$U_j R^{jj} = V_j,$$

where R^{jj} is the jth diagonal element of the inverse of the p by p predictor intercorrelation matrix. Hazewinkel (1963), in discussing the CV unique contributions, used R^{jj} and noted that it is the standard error of estimate for the regression of variable j onto the other $p-1$ predictors.

Issues With the Variance Partitioning Methods

One issue with all these variance partitioning methods is that they may lead to negative joint contributions due to computing them by subtraction, as pointed out by Bock (1975) and implied by Guilford (1965). These contributions obviously cannot be variance components (a negative variance would imply an imaginary variable). The geometry introduced in this article can be used to explain this issue.

The simplest geometric interpretation is that associated with U_1 and U_2 in Equation 5, and the basic geometry used here is explained in the Appendix. Assuming a standardized metric, with $\boldsymbol{z}_y$, $\boldsymbol{z}_1$, and $\boldsymbol{z}_2$ representing the dependent variate and predictors in this metric, the vector resultant of the projection of $\boldsymbol{z}_y$ onto $\boldsymbol{z}_2$, denoted as $\boldsymbol{p}_{\left(z_y \text{ on } z_2\right)}$, is $r_{y2}\boldsymbol{z}_2$. When the scalar, r_{y2}, is multiplied by the vector, $\boldsymbol{z}_2$, the result is a vector collinear with $\boldsymbol{z}_2$ but of length $||r_{y2}\boldsymbol{z}_2|| = r_{y2}||\boldsymbol{z}_2|| = r_{y2}$ (with $\boldsymbol{z}_2$ being in standard metric its length is 1.0, as described in the Appendix). The result of the projection of $\boldsymbol{z}_y$ onto $\boldsymbol{z}_2$ is identical to that of the projection of the vector, $\hat{\boldsymbol{z}}_y$ (values of Y predicted by the standardized regression equation), onto $\boldsymbol{z}_2$ and is shown in Figure 2, in which $\boldsymbol{v}_1$ is $\hat{\boldsymbol{z}}_y - \boldsymbol{p}_{\left(\hat{z}_y \text{ on } z_2\right)}$ (see Figures A2 and A3 and the accompanying text for more information about projections). Using Pythagoras's theorem, and referring to Figure 2 (the notation $\boldsymbol{v}_1 \perp \boldsymbol{z}_2$ indicates that

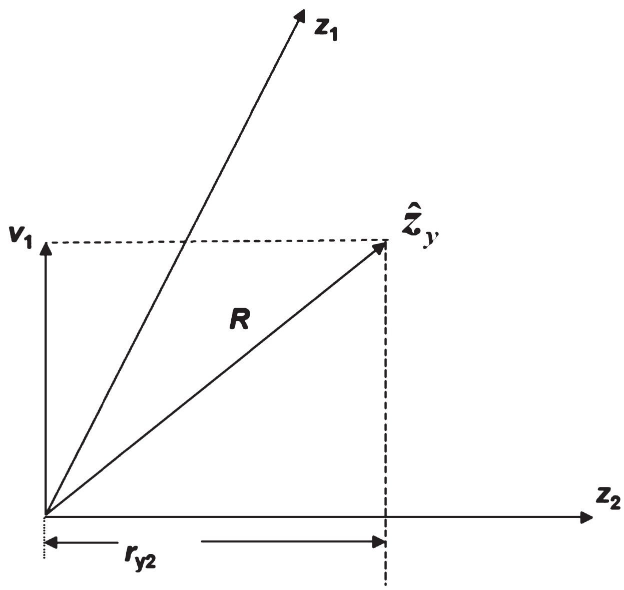

Figure 2 U_1 is the square of the length of v_1 defined as orthogonal to the projection of z_y onto z_2, with the vector resultants of the projection being $r_{y2}z_2$ (of length r_{y2}) and $v_1 \perp z_2$.

the two vectors are orthogonal, or perpendicular), it may be seen that (here I shorten the notation $R^2_{y.12}$ to R^2 to simplify the expressions),

$$\begin{aligned} \left\|\hat{\mathbf{z}}_y\right\|^2 = R^2 &= \left\|r_{y2}\mathbf{z}_2\right\|^2 + \left\|\mathbf{v}_1\right\|^2 \\ &= r^2_{y2} + \left\|\mathbf{v}_1\right\|^2, \end{aligned} \tag{7}$$

using Equation A4 in the final step. Thus from Equation 5:

$$U_1 = R^2 - r^2_{y2} = \left\|\mathbf{v}_1\right\|^2. \tag{8}$$

Note again that the projections of z_y and $\hat{z}_y$ onto z_2 are identical.

Note that this definition of the unique contribution of X_1, designated U_1, equals the squared length of a vector, v_1, that is orthogonal to (uncorrelated with) z_2 and has no direct relationship to the variable in question, X_1. By examining Figure 2 and Figure A3, it can be seen clearly that vectors $\hat{z}_y$, z_2, and v_1 lie in a plane that does not contain vector z_1. Hence my statement that U_1, the CV definition of the contribution of X_1, has no direct relationship to that variable. It is related to the linear regression model for the prediction of Y from X_2 independent of X_1 and can be considered as a contribution only in the sense that it represents the increase in predictable variation over and above that from the regression with X_2 as the only regressor.

Similarly, as shown in Figure 3, again using Equation 5,

$$U_2 = \left\|v_2\right\|^2. \tag{9}$$

So the unique contribution of X_2 by this definition is based on v_2, which is related to z_1 rather than z_2.

Thus the geometric interpretations of the two equations in Equation 5 are easy to see. If $\hat{z}_y$ is resolved into two orthogonal component vectors, v_1 and $r_{y2}z_2$, we have a decomposition of R^2 into the two orthogonal components U_1 and r^2_{y2}. A similar interpretation can be made from U_2 and r^2_{y1} with $v_2 \perp r_{y1}z_1$.

To this point, I have illustrated one problem with the CV definition of the contributions of regressors to the predicted variance in linear models. Geometrically, the definition of unique contribution of a predictor is the length of a vector not related to the regression involving that predictor. Next, I will outline Pappus's theorem and explain how it can be used to provide explanation of further problems with definitions of variance contributions.

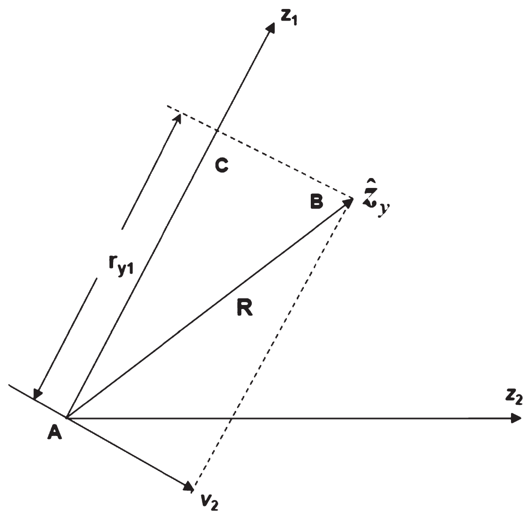

Figure 3 U_2 is the square of the length of $\mathbf{v}_2$ defined as orthogonal to the projection of $\mathbf{z}_y$ onto $\mathbf{z}_1$, with the vector resultants of the projection being the $r_{y1}\mathbf{z}_1$ (of length r_{y1}) and $\mathbf{v}_2 \perp \mathbf{z}_1$.

Pythagoras's and Pappus's Theorems

In this section, I discuss Pythagoras's theorem and its generalization to non-right triangles in a theorem of Pappus. Following those discussions, I return to the geometry, discussing how those theorems can be used for more complete understanding of the various definitions of contributions to explained variance discussed previously.

Pythagoras's Theorem: A Special Case of Pappus's Theorem and Relationship to Regression

To help introduce Pappus's theorem, I start this discussion by illustrating Pythagoras's theorem for right triangles as a special case of Pappus's theorem for any triangles. Pythagoras's theorem, of course, states that the squared length of the hypotenuse of a right triangle is equal to the sum of the squares of the lengths of the other two sides. Typically, the theorem is illustrated geometrically by constructing squares on the three sides, and the areas of the squares equal the three squared lengths in question. Pappus's theorem, on the other hand, deals with constructing parallelograms on the sides of any triangle, right or non-right. Recall that a square is a special case of a parallelogram; we use that fact in the following construction of Pythagoras's theorem as a special case of Pappus's theorem, which is discussed in the next section. Figure 4 illustrates the construction, using triangle ABC from Figure 3, representing part of the two-predictor regression discussed previously.

In Figures 3 and 4, ABC is a right triangle with the right angle at C. The triangle is formed with side AB that is the vector $\hat{\mathbf{z}}_y$ of predicted values on $\mathbf{z}_y$ in the two-predictor regression discussed above, and side AC that is the projection of $\hat{\mathbf{z}}_y$ (and hence, as pointed out previously, the $\mathbf{z}_y$ vector's projection as well) onto the vector, $\mathbf{z}_1$, of values of X_1. Pappus's theorem begins with construction of parallelograms on two sides of the triangle, so in this special case representing Pythagoras's theorem, we construct squares on sides AC (green square) and BC (blue square). The next step is to extend the sides of these two parallelograms to meet at P. Then we add two lines (AA′ and BB′) of the same length and in the same direction as PC, at A and B, respectively, and connect these two lines forming A′B′. We now have a third parallelogram on the third side, AB. Pappus's theorem states that the area of this parallelogram (AA′B′B) is equal to the sum of the areas of the parallelograms (AA″C′C and BB″C″C) constructed on the other two sides.

Again referring to Figure 3, note that the area of square AA′B′B is equal to the squared length of $\hat{z}_y$ the vector of predicted values on $\mathbf{z}_y$. Hence, as discussed previously (see Equation A6 and the two expressions immediately preceeding it), its area is equal to R^2, the proportion of variance accounted for by the two-predictor regression. Comparing Figures 2, 3, and 4, we can see that (a) the area of AA″C′C is equal to r_{y1}^2, the proportion of predicted variance in Y attributable to the one-predictor regression of Y onto X_1; and (b) the area of BB″C″C is the squared length of $\mathbf{v}_2$, hence equal to U_2, the CV contribution of X_2 in the two-predictor regression. A similar construction on the triangle in Figure 2 would show the geometry of the relationship between r_{y2}^2, the proportion of predicted variance in Y attributable to the one-predictor

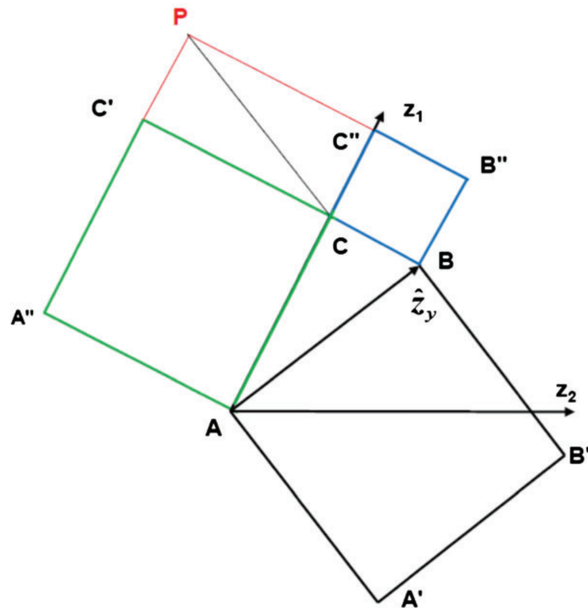

Figure 4 Pythagoras's theorem as a special case of Pappus's theorem.

regression of Y onto X_2, and the squared length of $\mathbf{v}_1$, equal to U_1, the CV contribution of X_1 in the two-predictor regression, respectively.

In the next section, I discuss the theorem of Pappus and how it is related to the issues of variance contributions in linear models.

Pappus's Theorem and Relationship to Regression

Kazarinoff (1961) presented what he described as "a little known but beautiful generalization of the Pythagorean Theorem, a theorem due to Pappus" (p. 84).[2] The theorem, as stated by Kazarinoff, follows ($\overline{\mathrm{AP}}$ denotes the length of line AP). Figure 5 illustrates the theorem.

> Let ABC be any triangle. Let AA′C′C and BB′C″C be any two parallelograms constructed on AC and BC so that either both parallelograms are outside the triangle or both are not entirely outside the triangle. . . . Prolong their sides A′C′ and B′C″ to meet in P. Construct a third parallelogram ABP″P′ on AB with AP′ and BP″ parallel to CP and with $\overline{\mathrm{AP'}} = \overline{\mathrm{BP''}} = \overline{\mathrm{CP}}$. The area of [parallelogram] ABP″P′ is equal to the sum of the areas of the parallelograms AA′C′C and BB′C″C.[3] (pp. 84–85)

The proof of the theorem is described by G. D. Allen (2000) as well as by Kazarinoff. Pythagoras's theorem can be used to illustrate the orthogonal partitioning of variation in linear statistical models. I will show how Pappus's theorem can be used in explanation of problems in partitioning variation in the nonorthogonal case.

There is a geometric interpretation of the CEG variance partitioning method, based on Equation A6, using Pappus's theorem with a non-right triangle. In Figure 6, I show the vectors $\mathbf{z}_1$, $\mathbf{z}_2$, and $\hat{\mathbf{z}}_y = b_1^*\mathbf{z}_1 + b_2^*\mathbf{z}_2$ in the plane onto which $\mathbf{z}_y$ was projected in Figure A3 (but using standard metric here). I also show the lengths $b_1^*||\mathbf{z}_1||$ and $b_2^*||\mathbf{z}_2||$ of the two vectors $b_1^*\mathbf{z}_1$ and $b_2^*\mathbf{z}_2$, multiples of $\mathbf{z}_1$ and $\mathbf{z}_2$, respectively, that are terms in the two-predictor regression equation.

The lengths of these vectors are $||b_1^*\mathbf{z}_1|| = b_1^*$ and $||b_2^*\mathbf{z}_2|| = b_2^*$ because $\mathbf{z}_1$ and $\mathbf{z}_2$ are standardized and thus have length 1.0.

Using triangle ABC from Figure 6, we next construct Figure 7 to show how Pappus's theorem applies. We form square ABB′A′ (red) having sides of length R and hence area equal to R^2, the variance accounted for by the two-predictor regression. We also construct squares, AA″C′C (green) and BB″C″C, (blue) on sides AC and BC, respectively, with sides of length b_1^* and b_2^* hence areas equal to $\left(b_1^*\right)^2$ and $\left(b_2^*\right)^2$. Thus the areas of these two squares equal the CEG unique variance contributions, V_1 and V_2, respectively, based on Equation 1.

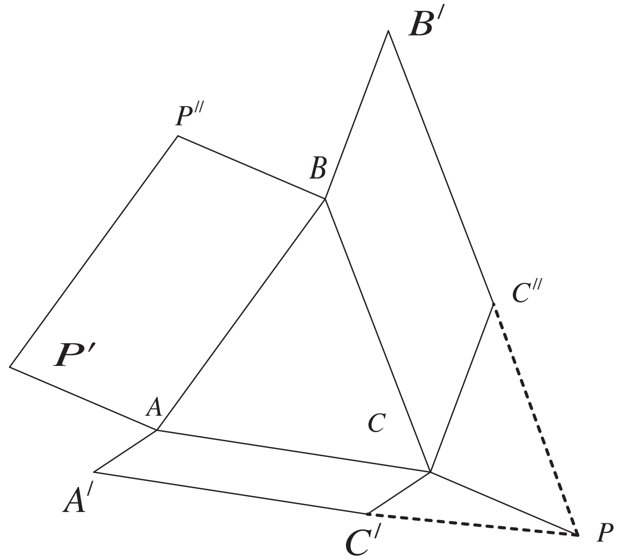

Figure 5 Diagram of Pappus's theorem.

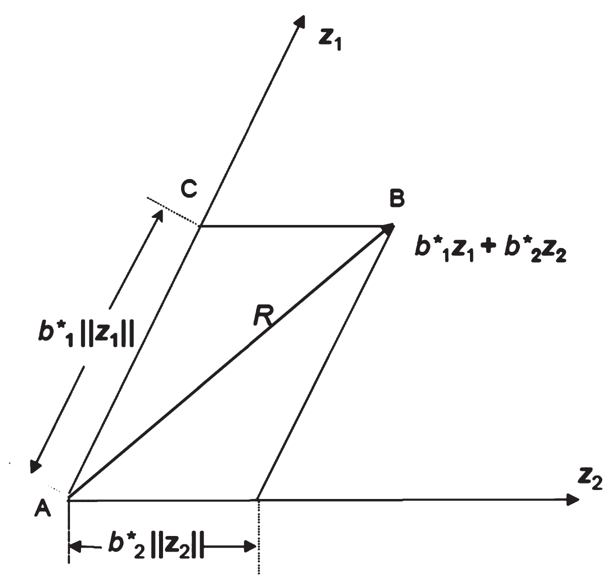

Figure 6 The plane spanned by $\mathbf{z}_1$ and $\mathbf{z}_2$, from Figure A3 but in standard metric.

From C, we then construct CP perpendicular to AB of length R. Next, we construct lines PQ and PS parallel to AC and BC, respectively. Extending the sides of square AA″C′C to meet PQ at M and N and similarly extending the sides of square BB″C″C to meet PS at M′ and N′ results in two rectangles, AMNC and CBN′M′, whose areas sum to the area of square AA′B′B by Pappus's theorem.

This means that the proportion of the total variance accounted for by the regression, R^2, can be partitioned into four components, two of which are squares representing the CEG unique contributions, $V_1 = \left(b_1^*\right)^2$ and $V_2 = \left(b_2^*\right)^2$. The remaining two rectangular portions, A″MNC′ and C″B″N′M′, must sum to the quantity $J_{12} = 2b_1^* b_2^* r_{12}$ from Equation 2. As mentioned previously, this quantity has been called the joint contribution of the two predictors in the CEG method.

Modifying Figure 7 and simplifying it so the figure is less cluttered, we form Figure 8, which will help to show the geometry of some of the other contributions. Referring to Figure 8, adding the dotted line from C to intersect side AB at a right angle at O, and letting angle CAB be α^{o}, it may be seen that angle ACO is of size $(90 - \alpha)^{\mathrm{o}}$ as is the angle between PC and $\mathbf{z}_1$. Hence angle PCN is also α^{o}. We saw previously that PC is of length R, hence NC is of length $R\ \cos(\alpha^{\mathrm{o}})$.

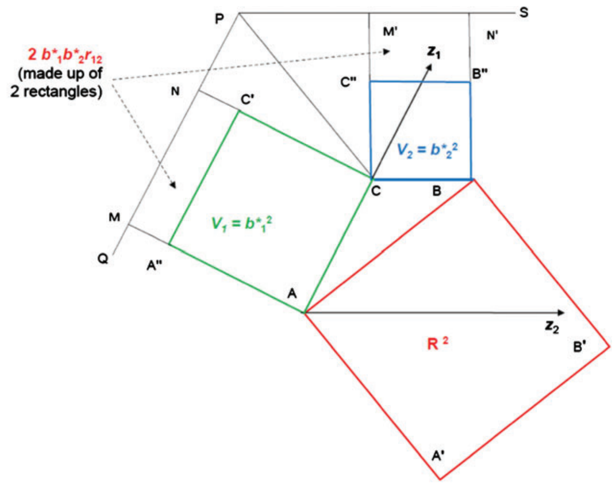

Figure 7 Pappus's theorem illustrating Chase-Engelhart-Guilford (CEG) contributions.

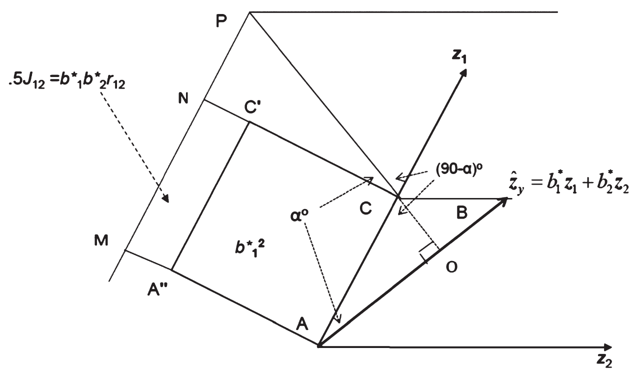

Figure 8 Figure 7 modified to show angles and additional areas.

Recalling that cosines of angles between vectors are correlations, from the figure we can see that this cosine is equal to the correlation, $r_{\hat{\mathbf{z}}_y \mathbf{z}_1}$. Recall that correlations are invariant over linear transformations so the metric is irrelevant; however, for consistency, I continue to use the standard metric. This correlation is the ratio of the covariance between the two variables to the product of their standard deviations,

$$r_{\mathbf{z}_1 \hat{\mathbf{z}}_y} = \frac{s_{\mathbf{z}_1 \hat{\mathbf{z}}_y}}{s_{\mathbf{z}_1} s_{\hat{\mathbf{z}}_y}}.$$

because we are using standard metric, $s_{\mathbf{z}_1} = 1$. Furthermore,

$$\begin{aligned} s_{\mathbf{z}_1 \hat{\mathbf{z}}_\mathrm{y}} &= \mathbf{z}_1' \hat{\mathbf{z}}_\mathrm{y} = \mathbf{z}_1' \left(b_1^* \mathbf{z}_1 + b_2^* \mathbf{z}_2 \right) \\ &= b_1^* \mathbf{z}_1' \mathbf{z}_1 + b_2^* \mathbf{z}_1' \mathbf{z}_2 \\ &= b_1^* + b_2^* r_{12}. \end{aligned}$$

and $s_{\hat{\mathbf{z}}_y} = R$. Hence,

$$\cos\left(\alpha^{\mathrm{o}}\right) = r_{\mathbf{z}_1 \hat{\mathbf{z}}_y} = \frac{b_1^* + b_2^* r_{12}}{R},$$

and the length of NC is:

$$R \cos\left(\alpha^{\mathrm{o}}\right) = b_1^* + b_2^* r_{12}.$$

Because CC′ is of length b_1^*, NC′ is of length $b_2^* r_{12}$ and rectangle A″C′NM has area $b_1^* b_2^* r_{12}$. Similarly it may be seen that rectangle B″N″M′C″ has this same area, $b_1^* b_2^* r_{12}$. Thus, as stated previously and illustrated in Figure 7, the areas of

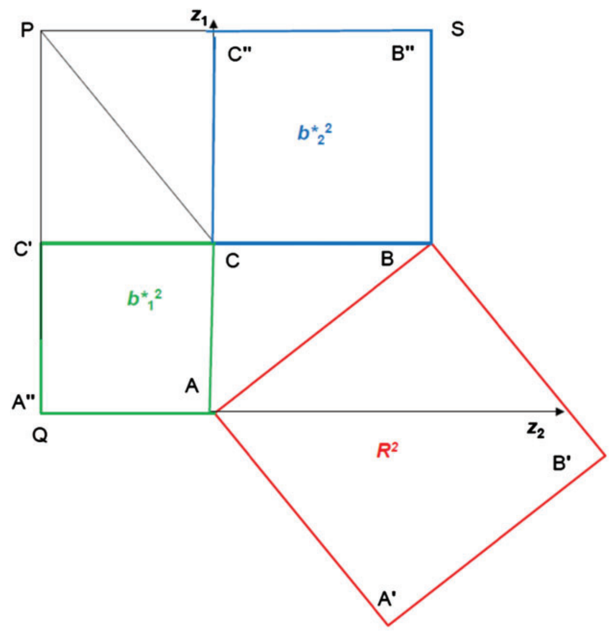

Figure 9 Chase-Engelhart contributions in the orthogonal case: Pythagoras's theorem.

these two rectangles sum to the Chase-Engelhart joint contribution, $J_{12} = 2b_1^* b_2^* r_{12}$: and each of these two rectangles is of the same area, $J_{12}/2$.

Hence these two individual rectangles do not each equal Guilford's (1965) indirect contributions discussed above. Rather each equals one-half of the CV joint contribution. From the geometry, the author concludes that problems exist with all of these definitions of contributions of predictors to the overall regression.

I will now show additional reasons why the variance partitioning procedures under discussion are not appropriate. Consider moving point C in Figure 7 farther from AB (also moving P because CP must be parallel to and equal in length to AA′ by the theorem) as would occur if X_1 and X_2 were correlated lower than in our example. We first move them far enough that angle ACB becomes a right angle (Figure 9). The squares representing $\left(b_1^*\right)^2$ and $\left(b_1^*\right)^2$ extend to lines PQ and PS, eliminating the two rectangular areas that summed to $2b_1^* b_2^* r_{12}$. Hence when r_{12} is zero, this joint contribution, and Guilford's (1965) indirect contributions, are zero. Thus the accounted for variance is broken down into two components that can be uniquely associated with X_1 and X_2, respectively. This is the orthogonal case, which is also the only case in which there is a single unambiguous definition of the unique contributions of each predictor (Carlson, 1968). Note that this is the special case of Pappus's theorem that is the old familiar Pythagoras's theorem: the square on the hypotenuse is equal to the sum of the squares on the other two sides.

The real concern, however, is when the correlation between the two predictors is negative, so C is moved even farther out as shown in Figure 10. And note that there is no reason that negatively correlated predictors should not be used in a regression analysis. In Figure 10, I left P, Q , and S in locations such that PC is still the same length (R) as AA′ and BB′. This allows me to show that the sum of the CEG unique contributions, $V_1 = \left(b_1^*\right)^2$ and $V_2 = \left(b_2^*\right)^2$, is greater than R^2, the total variance accounted for by the regression. That is, the sum of the areas of squares AA″C′C and BB″C″C is greater than the area of square ABB′A′. Thus, treating the Chase-Engelhart contributions as representing the variance contributions of the two predictors is inappropriate; they account for more than the total. In this case, the right-hand side of Equation 2 is negative and obviously cannot be interpreted as a variance component. As previously mentioned, however, Guilford (1965) stated that for interpretation of any product of a standardized regression coefficient, b^*, with a correlation, r, as a contribution, that product must be positive, and Bock (1975), more correctly stated that it must be nonnegative.

Discussion

In this article, I demonstrated that the various components of variation that have sometimes been used in conjunction with least squares estimation using linear statistical models having two independent variables and one dependent variate

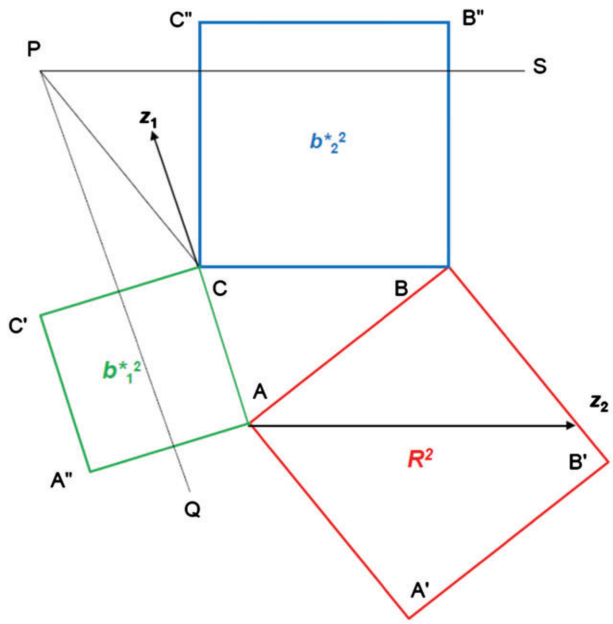

Figure 10 Chase-Engelhart contributions with negatively correlated predictors.

are interpretable as areas in a plane that is a subspace of an N-dimensional space. As a result, one can see geometrically why the various definitions of contributions to variation sometimes lead to negative quantities that are not interpretable as contributions to the variation accounted for by the linear model. Although the presentation has been in terms of the regression model, the situation is identical for a simple analysis of variance model having only two columns in the design matrix (and assuming that it is formulated with a full-rank design matrix; Speed, 1969; Timm & Carlson, 1975).

All of this leads to the fact that the only situation in which the accounted for variation can be unambiguously partitioned into portions uniquely associated with each independent variable is the case of orthogonal regressors, or orthogonal designs in the analysis of variance context. In the analysis of variance (ANOVA) context, the different partitions can be shown to relate to different hypothesis tests for nonorthogonal designs (Searle, 1971; Speed, 1969; Timm & Carlson, 1975), a fact not well known by users of popular computer programs that allows the user to compute the different partitions.[4] The areas used in the explanations in this article relate to the squared lengths of vectors, quantities that can only be summed meaningfully under orthogonality, in which case Pythagoras's theorem applies. We are dealing with the projection of the $\mathbf{y}$ vector onto the plane spanned by two linearly independent $\mathbf{x}$ vectors, and we thus are limited to resolving the vector resulting from the projection (the $\hat{\mathbf{y}}$ vector) into two linearly independent component vectors because any third vector in the plane must be linearly dependent on the first two (we can, of course, resolve $\hat{\mathbf{y}}$ into more than two linearly dependent but nonorthogonal components). All of these notions easily extend to the more general case of p independent variables (assuming, of course, that the p independent variables are a linearly independent set). It is harder to represent the more general case on paper because the predictor space is a hyperplane (of more than two dimensions). When we use the p predictor linear regression model, it is possible to resolve the accounted for variation into p orthogonal components, but interpretation of each of these components as uniquely associated with one of the predictors is possible only when the p vectors are all orthogonal to one another. The use of stepwise regression, or alternative one-at-a-time entry of regressors into a regression equation, and accompanying designation of contributions of each of the p predictors suffer from the same fallacy.

Another interesting case is that in which there are two or more sets of $\mathbf{x}$ vectors such that all vectors within any one set are orthogonal to all vectors not in that set but not necessarily orthogonal to vectors within the same set. This is the case in orthogonal factorial analysis of variance models, and in this case, the accounted for variation can be partitioned into portions uniquely associated with the main effects and interaction effects of factors.

It is always possible to resolve $\hat{\mathbf{y}}$ into p orthogonal components given p linearly independent $\mathbf{x}$ vectors (whether they are orthogonal or not), but an infinite number of different resolutions exist. A unified way to look at these different resolutions is through the vehicle of factoring the matrix of correlations of the X variables using different orthogonal factoring

techniques and regressing Y onto the resulting orthogonal variables. As was mentioned earlier in this report the use of U_j relates to Cholesky factoring of the matrix. The reason that the U_j are not independent (unless the $\mathbf{x}_j$ are orthogonal) is that each is based upon a different factoring (beginning the decomposition with a different variable) rather than all of them being based on one orthogonal factoring. The practice of defining U_j as the unique contribution of X_j in the nonorthogonal case can be seen, using Pappus's theorem similarly to its use previously, to be no more defensible than any of several other definitions. Recalling that U_1 is the squared length of $\mathbf{v}_1$ in Figure 2 and U_2 the squared length of $\mathbf{v}_2$ in Figure 3, one can question the legitimacy of referring to the implied variation partitioning as partitioning into two components uniquely associated with X_1 and X_2. One could argue that it would be equally legitimate to resolve $\hat{\mathbf{y}}$ into two orthogonal vectors, $\mathbf{w}_1$ and $\mathbf{w}_2$, such that the angle between $\mathbf{w}_1$ and $\mathbf{x}_1$ is the same as that between $\mathbf{w}_2$ and $\mathbf{x}_2$. This procedure involves nothing more than the regression of Y onto the principal components of X_1 and X_2 rotated through a 45° angle. That there are an infinite number of different orthogonal factorings that could be used can be seen by considering all possible right triangles constructed within a semicircle. The situation is more complex when there are more than two X variables. Several different factoring techniques have been suggested and used in conjunction with regression analysis (Burket, 1964; Creager, 1969; Creager & Boruch, 1969; Darlington, 1968; Horst, 1941; Lawley & Maxwell, 1973; Massey, 1965; Reed, 1941).

The geometry presented in this article can also be used to illustrate issues dealing with variance contributions in the case of subtests that are combined into an overall total test score. The issues of the contributions of subtest score variances to the variance of the composite score parallel those of the contributions of regressors to the prediction of a dependent variable discussed in this article (see Carlson, in press). Further, the effects of differential weighting of subtests before summing to obtain the composite measure can also be illustrated with the geometry.

Noting that all the techniques mentioned in this article are based on least squares, it seems legitimate to raise the question of why this procedure should be used in estimation. It does, of course, provide the estimator with minimum variance, among all linear unbiased estimators, and the mathematics involved are relatively simple. In the case of gross departures from orthogonality, however, it is well known in the statistical community that least squares yields estimators of regression coefficients that have such large variances that any of several other techniques that yield biased but much less variable estimators is better, as far as estimation is concerned (D. M. Allen, 1974; Andrews, 1974; Hocking, Speed, & Lynn, 1975; Hoerl & Kennard, 1970; Marquardt, 1970; Marquardt & Snee, 1975; Webster, Gunst, & Mason, 1974). The relevance to the current topic is that the variance partitioning schemes discussed here are based on these estimates that could be unstable under collinearity.

Notes

1 Independence, of course, is a broader term than zero correlation; the latter refers only to linear independence.
2 The theorem and its proof are also presented in G. D. Allen (2000, p. 4).
3 For our purposes we need to only consider the case of parallelograms outside the triangle.
4 The statistics associated with these different hypotheses are sometimes referred to as Type I, Type II, and Type III sums of squares. The references cited show that using the different types of sums of squares involves testing different hypotheses. These hypotheses include sample sizes as weights and hence are only reasonable in the case of sample sizes being proportional to population sizes, a fact that users appear to ignore or to be unaware of.

References

Allen, D. M. (1974). The relationship between variable selection and data augmentation and a method for prediction. *Technometrics*, *16,* 125–127.

Allen, G. D. (2000). *Pappus of Alexandria (fl. c. 300–c. 350)*. Retrieved from http://www.math.tamu.edu/~dallen/masters/Greek/pappus.pdf

Andrews, D. F. (1974). A robust method for multiple linear regression. *Technometrics*, *16,* 523–531.

Bancroft, T. A. (1968). *Topics in intermediate statistical methods*. Ames: Iowa State University Press.

Bock, R. D. (1975). *Multivariate statistical methods in behavioral research*. New York, NY: McGraw-Hill.

Burket, G. R. (1964). A study of reduced rank models for multiple prediction. *Psychometric Monograph No. 12*.

Carlson, J. E. (1968). *Effect of differential weighting on the interrater reliability of essay grades*. Unpublished doctoral dissertation, University of Alberta, Edmonton, Canada.

Carlson, J. E. (in press). *What differential weighting of subsets of items does and does not accomplish: Geometric explanation* (Research Report). Princeton, NJ: Educational Testing Service.

Carlson, J. E., & Carlson, J. V. (2008a, March). *A generalization of Pythagoras' theorem and application to explanations of variance contributions in linear models.* Paper presented at the meeting of the American Educational Research Association, New York, NY.

Carlson, J. E., & Carlson, J. V. (2008b, March). *What differential weighting of subsets of items does and does not accomplish: Geometric explanation*. Paper presented at the meeting of the National Council on Measurement in Education, New York, NY.

Carlson, J. E. & Timm, N. H. (1974). Analysis of nonorthogonal fixed effects designs. *Psychological Bulletin*, *81,* 63 – 70.

Chase, C. I. (1960). Computation of variance accounted for in multiple correlation. *The Journal of Experimental Education*, *28,* 265 – 266.

Creager, J. A. (1969). The interpretation of multiple regression using overlapping rings. *American Educational Research Journal*, *6,* 706 – 709.

Creager, J. A. (1971). Orthogonal and nonorthogonal methods for partitioning regression variance. *American Educational Research Journal*, *8,* 671 – 676.

Creager, J. A., & Boruch, R. F. (1969, August). *Orthogonal analysis of linear composite variance*. Paper presented at the 77th convention of the American Psychological Association, Washington, DC.

Creager, J. A. & Valentine, L. D. (1962). Regression analysis of linear composite variance. *Psychometrika*, *27,* 31 – 37.

Darlington, R. B. (1968). Multiple regression in psychological research and practice. *Psychological Bulletin*, *69,* 161 – 182.

Draper, N. R. & Smith, H. (1966). *Applied regression analysis*. New York, NY: Wiley.

Engelhart, M. D. (1936). The technique of path coefficients. *Psychometrika*, *1,* 287 – 293.

Guilford, J. P. (1965). *Fundamental statistics in psychology and education*. New York, NY: McGraw-Hill.

Hazewinkel, A. (1963). A note concerning the Creager Valentine paper. *Psychometrika*, *28,* 105 – 108.

Hocking, R. R., Speed, F. M., & Lynn, J. J. (1975). *A class of biased estimators in linear regression*. Starkville: Mississippi State University, Department of Computer Science and Statistics.

Hoerl, A. E. & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12,* 55 – 67.

Horst, P. (1941). Mathematical contributions: Supplementary study E. In P. Horst (Ed.), *The prediction of personal adjustment* New York, NY: Social Science Research Council.

Kazarinoff, N. D. (1961). *Geometric inequalities*. New York, NY: Random House.

Kendall, M. G. & Stuart, A. (1967). *The advanced theory of statistics* (Vol. *2,* 2nd ed.). London, England: P. Griffin.

Lawley, D. N. & Maxwell, A. E. (1973). Regression and factor analysis. *Biometrika*, *60,* 331 – 338.

Marquardt, D. W. (1970). Generalized inverses, ridge regression, biased linear estimation and nonlinear estimation. *Technometrics*, *12,* 591 – 612.

Marquardt, D. W. & Snee, R. D. (1975). Ridge regression in practice. *The American Statistician*, *29,* 3 – 20.

Massey, W. F. (1965). Principal components regression in exploratory statistical research. *Journal of the American Statistical Association*, *60,* 234 – 256.

Mood, A. M. (1971). Partitioning variance in multiple regression as a tool for developing learning models. *American Educational Research Journal*, *8,* 191 – 202.

Newton, A. G. & Spurrell, D. J. (1967). A development of multiple regression for the analysis of routine data. *Applied Statistics*, *16,* 51 – 64.

Pugh, R. C. (1968). The partitioning of criterion score variance accounted for in multiple regression. *American Educational Research Journal*, *5,* 639 – 646.

Reed, R. B. (1941). An empirical study in the reduction of the number of variables used in prediction: Supplementary study C. In P. Horst (Ed.), *The prediction of personal adjustment*. Social Science Research Council: New York, NY.

Richardson, M. W. (1941). The combination of measures: Supplementary study D. In P. Horst (Ed.), *The prediction of personal adjustment*. Social Science Research Council: New York, NY.

Searle, S. R. (1971). *Linear models*. New York, NY: Wiley.

Speed, F. M. (1969). *A new approach to the analysis of linear models* (NASA Technical Memorandum TM X-58030). Houston, TX: National Aeronautics and Space Administration.

Timm, N. H. (1975). *Multivariate analysis with applications in education and psychology*. Belmont, CA: Brooks-Cole.

Timm, N. H., & Carlson, J. E. (1975). Analysis of variance through full rank models. *Multivariate Behavioral Research Monograph No. 75-1*.

Webster, J. J., Gunst, R. F., & Mason, R. L. (1974). Latent root regression analysis. *Technometrics*, *16,* 513 – 522.

Wickens, T. D. (1995). *The geometry of multivariate statistics*. Hillsdale, NJ: Erlbaum.

Wisler, C. (1968). *On partitioning the explained variance in a regression analysis* (Technical Note No. 2). Washington, DC: US Office of Education.

Wonnacott, R. J. & Wonnacott, T. H. (1973). *Econometrics*. New York, NY: Wiley.

Yates, F. (1934). The analysis of multiple classifications with unequal numbers in the different classes. *Journal of the American Statistical Association*, *29,* 51–66.

Appendix

Vector Algebra and Geometry Related to Regression

Addition of Vectors

I illustrate in two dimensions with vector $\mathbf{X}_1$, a column vector, having values of 0 and 3, indicating that there are two persons who had values of 0 and 3 and, similarly, having values of 1.5 and 2 on vector $\mathbf{X}_2$. Vector addition, of course, is an element-by-element operation, in this case,

$$\mathbf{X}_1 + \mathbf{X}_2 = \begin{pmatrix} 0 \\ 3 \end{pmatrix} + \begin{pmatrix} 1.5 \\ 2 \end{pmatrix} = \begin{pmatrix} 1.5 \\ 5 \end{pmatrix} = \mathbf{X}_S,$$

where $\mathbf{X}_S$ represents the sum. I note for future reference that the transpose of vector $\mathbf{X}_1$ is defined as the row vector,

$$\mathbf{X}_1' = \begin{pmatrix} 0 & 3 \end{pmatrix}.$$

To show the geometry of vector addition, in Figure A1 I show $\mathbf{X}_1$ as a vector from the origin to the point having coordinates (0, 3) and, similarly, $\mathbf{X}_2$ as a vector from the origin to (1.5, 2). A copy of $\mathbf{X}_2$ is appended to the end of $\mathbf{X}_1$ as a vector from (0, 3) to (1.5, 5). As shown in a number of sources (e.g., Wickens, 1995, p. 11), the sum $\mathbf{X}_1 + \mathbf{X}_2$ is a vector from the origin (0, 0) to the end of the appended $\mathbf{X}_2$ vector with coordinates (1.5, 5). In this case, it represents the addition of two variables to form a variable sum, in vector $\mathbf{X}_S$. Using Pythagoras's theorem, it is easily seen that the length of a vector is the square root of the sum of the squared elements.

Other Geometric Properties of Vectors

Using $\mathbf{X}_2$ in Figure A1 as an example of sample data, it is easily seen using Pythagoras's theorem that the length of the vector is the square root of the sum of the squares of the values in the vector (Wickens, 1995, p. 10). If the variables in a sample are transformed to deviation scores (subtract the mean from each score), the length becomes the square root of the deviation sum of squares, which if divided by $\sqrt{N-1}$ will equal the variables' standard deviation estimates (Wickens, 1995, p. 19). From this point on I denote the deviation score vectors as $\mathbf{x}_1$ and $\mathbf{x}_2$. As pointed out by Wickens (p. 19), "In most analyses, the constant of proportionality $\sqrt{N-1}$ is unimportant, since every vector is based on the same number of observations. One can treat the length of a vector as equal to the standard deviation of its variable."

Hence in the body of this report the lengths of vectors in the deviation score metric are usually described as equal to the standard deviations.

In addition, the cosine of the angle between two vectors of deviation scores equals the estimate of the Pearson correlation between the variables (Wickens, 1995, pp. 19–20). In the extremes, two collinear vectors (along the same line) represent variables that are correlated 1.0 (if in same direction) or −1.0 (if in opposite direction), and orthogonal vectors (at right angles) represent variables correlated 0.0.

An important part of the geometry of estimation in linear statistical modeling is the perpendicular projection of one vector onto another. As an illustration in Figure A2, I show the perpendicular projection of vector $\mathbf{y}$ onto vector $\mathbf{x}$ (a line is constructed from the end of vector $\mathbf{y}$ to intersect vector $\mathbf{x}$ at a right angle). The vector from the origin to the point of the projection on $\mathbf{x}$ is denoted as $\mathbf{p}_{(\mathbf{y}\text{ on }\mathbf{x})}$ and the length of the projection as $p_{(y\text{ on }x)}$.

We note also that subtracting a vector from another vector simply involves taking the negative of the second vector and adding it to the first (Wickens, 1995, p. 11). The result of subtracting from $\mathbf{y}$ the vector resulting from the projection in Figure A2 is the vector denoted as $\mathbf{y} - \mathbf{p}_{(\mathbf{y}\text{ on }\mathbf{x})}$.

The Unit and Deviation Score Vectors

An important vector in this geometry is the unit vector ($\mathbf{1}$, a vector of N 1s). The perpendicular projection of a vector, $\mathbf{Y}$ (recall this vector is in raw score units) onto the unit vector, $\mathbf{p}_{(\mathbf{y}\text{ on }\mathbf{1})}$ results in the mean vector, having all N elements equal

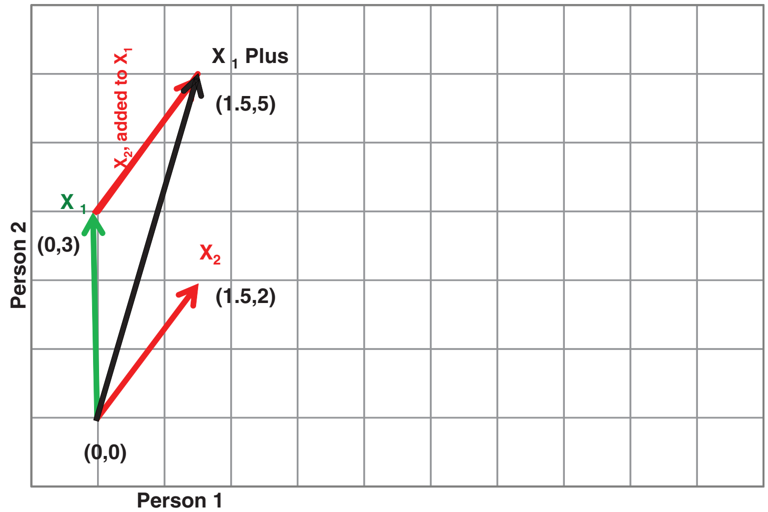

Figure A1 Addition of vectors.

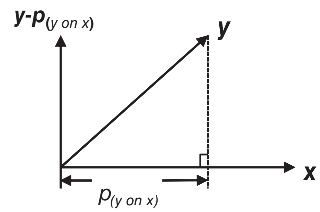

Figure A2 Projection of y onto x.

to the mean of **Y**. The difference between **Y** and the mean vector is the vector of deviation scores, **y**, all elements being deviations from the mean of **Y**. In other words, transforming sample values on a variable from raw scores to deviation scores (deviations from the mean) involves projecting the raw score vector onto two orthogonal components: one along the one-dimensional space defined by the unit vector, **1**, and the other (the deviation vector) in the $N-1$-dimensional subspace orthogonal to **1** (Wickens, 1995, pp. 37–38). We say we are decomposing the vector into two components, and it may be thought of as the reverse operation of the addition of two vectors to form a composite. We note for completeness that the degrees of freedom associated with a variance estimate, $N-1$, are equal to this dimensionality; we only need $N-1$ points in the space to specifically locate the vector, hence $N-1$ degrees of freedom. The N th deviation value, of course, can always be calculated from the other $N-1$ because deviation scores sum to zero. As shown by Wickens (p. 19) and mentioned previously, the deviation vector has length equal to the square root of the deviation sum of squares and is therefore proportional to the variables' standard deviation. Dividing the vector by the square root of the dimensionality ($\sqrt{N-1}$) results in a vector having length equal to the standard deviation. Further dividing the elements of the vector by the standard deviation is equivalent to transforming to standard scores (mean 0.0, standard deviation 1.0), and the resulting vector is of length 1.0, the standard deviation of a standardized variable.

Simple Linear Regression

Figure A2 can also be used to represent the geometry of OLS regression with one dependent variate and one independent variable, with N sample values (in deviation score metric) represented by the vectors **y** and **x**, respectively. The vector resulting from the projection is $\hat{\mathbf{y}} = \mathbf{p}_{(\mathrm{y\ on\ x})}$, the vector of predicted values on Y, and the difference vector is the vector of errors of prediction, $\mathbf{y} - \mathbf{p}_{(\mathrm{y\ on\ x})} = \mathbf{e}$.

From this point, we can consider the vectors independently of the number of persons defining the Euclidean space because two (linearly independent) vectors and their vector sum are always coplanar (embedded in a plane that is a two-dimensional subspace of the N-dimensional space, Wickens, 1995, pp. 25–28). That is, with a sample of size N the vectors

each have N elements and exist in an N-dimensional space. But the three vectors representing the simple linear regression all lie within a plane within that N-dimensional space as shown in Figure A2. Thus, the geometry we are interested in for our purposes can be represented in exactly the same way for two dimensions or N dimensions.

Correlation and Covariance

As mentioned previously and shown by Wickens (1995), the correlation coefficient is equal to the cosine of the angle between the deviation score vectors representing two variables in a sample (Wickens, 1995, pp. 19–20). This helps in understanding linear relationships between variables. Note that if the two variables were correlated 0.0 with each other, they would be represented as orthogonal (at right angles) vectors because the cosine of 90 ̂ is 0.0. The variance of their sum would be the simple sum of the two individual variables' variances, because, algebraically, the variance of the sum of two variables is well known to be

$$s_S^2 = s_1^2 + s_2^2 + 2s_1s_2r_{12}, \tag{A1}$$

where s_S^2 and s_1^2, s_2^2 are the variances of the sum and the two component variables, respectively; r_{12} is the correlation between the two component variables; and the term on the right, $2s_1s_2r_{12}$. is twice the covariance between the two component variables. Thus, if the two components correlate 0.0, the covariance is 0.0 and the composite variance is the sum of the two individual variances.

Regression With Two Regressors

Consider the regression of a dependent variate, Y, onto two linearly independent variables, X_1 and X_2. Given N observations in a sample, we may represent the values on Y, $\mathbf{X}_1$, and $\mathbf{X}_2$ as the three N-element vectors $\mathbf{Y}, \mathbf{X}_1$, and $\mathbf{X}_2$ in an N-dimensional space. If the values are scaled to standard form, as discussed previously, I will denote them as $\mathbf{z}_y$, $\mathbf{z}_1$, and $\mathbf{z}_2$ and these three vectors are all of unit length and lie within an $N-1$ dimensional subspace. Letting $\|\mathbf{z}_y\|$ denote the length of vector $\mathbf{z}_y$, several relationships may be derived (Timm, 1975; Wickens, 1995; Wonnacott & Wonnacott, 1973). The scalar (also called dot) product, defined as the sum of squared products of the elements, of any two of these standardized vectors is equal to the product–moment correlation coefficient between the two variables. For example,

$$\mathbf{z}_y'\mathbf{z}_1 = \sum_{i=1}^{N} z_{yi}z_{1i} = r_{y1}.$$

The length of a vector is the square root of the scalar product of the vector with itself and, as mentioned previously, is equal to 1.0 in standard metric:

$$\|\mathbf{z}_1\| = \left(\mathbf{z}_1'\mathbf{z}_1\right)^{1/2} = \left(\sum_{i=1}^{N} \mathbf{z}_{1i}^2\right)^{1/2} = 1.0. \tag{A2}$$

In this metric, the vector resulting from the perpendicular projection of $\mathbf{z}_y$ onto $\mathbf{z}_1$ is

$$\left(\mathbf{z}_y'\mathbf{z}_1\right)\mathbf{z}_1 = \left(\mathbf{z}_1'\mathbf{z}_y\right)\mathbf{z}_1 = r_{y1}\mathbf{z}_1, \tag{A3}$$

with squared length (recalling that the length of a vector in standard metric is 1),

$$\left\|r_{y1}\mathbf{z}_1\right\|^2 = r_{y1}^2\left\|\mathbf{z}_1\right\|^2 = r_{y1}^2. \tag{A4}$$

Equations A3 and A4 relate to the geometry of bivariate regression in the standardized metric and can be derived with trigonometry or by using Pythagoras's theorem (Wickens, 1995).

Regressing Y onto two independent variables, X_1 and X_2, using deviation score metric, geometrically is simply the perpendicular projection of $\mathbf{y}$ onto the plane spanned by $\mathbf{x}_1$ and $\mathbf{x}_2$ (Figure A3) when we use OLS estimation.

By "the plane spanned by $\mathbf{x}_1$ and $\mathbf{x}_2$," we refer to the fact that any two (linearly independent) vectors in an N-dimensional space are always enclosed in a plane that is a two-dimensional subspace of the N-dimensional space, and any other vector in that plane can be represented as a linear combination of those two. Any two such linearly independent vectors are referred to as *basis vectors* of the plane. If we use standardized variables, letting b_1^* and b_2^* represent the

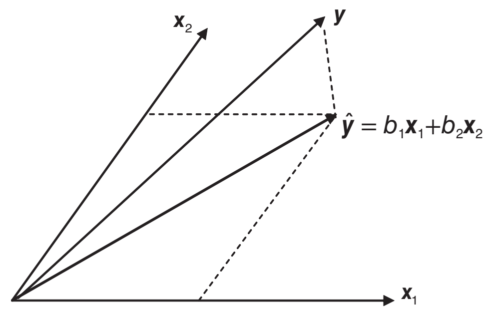

Figure A3 Regression: perpendicular projection of y onto the space spanned by x_1 and x_2.

(standardized) regression coefficients, R^2, the coefficient of determination (square of the multiple correlation, R), and $\widehat{\mathbf{z}}_y$, the vector resulting from the projection of $\mathbf{z}_y$ onto the plane, we may further develop the following relationships. The resultant vector of the projection is the vector of predicted values on the dependent variable and is the linear combination,

$$\widehat{\mathbf{z}}_y = b_1^* \mathbf{z}_1 + b_2^* \mathbf{z}_2. \tag{A5}$$

The multiple correlation coefficient, R, is the correlation between Y and $\widehat{Y}$, and as shown by Wickens (1995, p. 36) using Pythagoras's theorem, is the ratio of the length of $\widehat{\mathbf{y}}$ to that of $\mathbf{y}$. Here, because we are working in the standardized metric in which the length of $\mathbf{z}_y$ is one, the length of the projection is simply the length of $\widehat{\mathbf{z}}_y$. The squared length of $\widehat{\mathbf{z}}_y$ is (again recalling the $\mathbf{z}_1$ and $\mathbf{z}_2$ vectors have length one),

$$\begin{aligned}\left\|\mathbf{z}_{\hat{y}}\right\|^2 &= \left(b_1^*\right)^2 ||\mathbf{z}_1||^2 + \left(b_2^*\right)^2 ||\mathbf{z}_2||^2 + 2b_1^* b_2^* ||\mathbf{z}_1|| \cdot ||\mathbf{z}_2|| r_{12} \\ &= \left(b_1^*\right)^2 + \left(b_2^*\right)^2 + 2b_1^* b_2^* r_{12} = R^2,\end{aligned}$$

which can also be expressed as (Guilford, 1965, p. 398),

$$\left\|\widehat{\mathbf{z}}_y\right\|^2 = b_1^* r_{y1} + b_2^* r_{y2} = R^2.$$

Generalizations of these two formulas to the case of p regressors results in

$$R^2 = \sum_{j=1}^{p} \left(b_j^*\right)^2 + \sum_{j=1}^{p} \sum_{j' \neq j=1}^{p} 2b_j^* b_{j'}^* r_{jj'}, \tag{A6}$$

and

$$R^2 = \sum_{j=1}^{p} b_j^* r_{yj}. \tag{A7}$$

These expressions are used in the text of this article to discuss one method used to define the unique contributions of regressors to the prediction.

Suggested citation:

Carlson, J. E. (2014). *A generalization of Pythagoras's Theoremand applications to explanations of variance contributions in linear models* (Research Report No. RR-14-18). Princeton, NJ: Educational Testing Service. doi: 10.1002/ets2.12018

Action Editor: Shelby Haberman

Reviewers: J. R. Lockwood and John Donoghue